

On Sharp Performance Bounds for Robust Sparse Signal Recoveries

Weiyu Xu

Department of Electrical Engineering
California Institute of Technology
Pasadena CA 91125, USA
Email: weiyu@systems.caltech.edu

Babak Hassibi

Department of Electrical Engineering
California Institute of Technology
Pasadena CA 91125, USA
Email: hassibi@systems.caltech.edu

Abstract—It is well known in compressive sensing that l_1 minimization can recover the sparsest solution for a large class of underdetermined systems of linear equations, provided the signal is sufficiently sparse. In this paper, we compute sharp performance bounds for several different notions of robustness in sparse signal recovery via l_1 minimization. In particular, we determine necessary and sufficient conditions for the measurement matrix A under which l_1 minimization guarantees the robustness of sparse signal recovery in the “weak”, “sectional” and “strong” senses (e.g., robustness for “almost all” approximately sparse signals, or instead for “all” approximately sparse signals). Based on these characterizations, we are able to compute sharp performance bounds on the tradeoff between signal sparsity and signal recovery robustness in these various senses. Our results are based on a high-dimensional geometrical analysis of the null-space of the measurement matrix A . These results generalize the thresholds results for purely sparse signals [1], [3] and also present generalized insights on l_1 minimization for recovering purely sparse signals from a null-space perspective.

Index Terms: compressed sensing, basis pursuit, Grassmann angle, geometric probability, random linear subspaces

I. INTRODUCTION

In this paper we are interested in compressed sensing problems. Namely, we would like to find \mathbf{x} such that

$$A\mathbf{x} = \mathbf{y} \quad (1)$$

where A is an $m \times n$ ($m < n$) measurement matrix, \mathbf{y} is a $m \times 1$ measurement vector and \mathbf{x} is an $n \times 1$ unknown vector. In the conventional compressed sensing context, \mathbf{x} has only k ($k < m$) nonzero components. In this paper we will consider a more general version, namely, we will assume that k components of the vector \mathbf{x} have large magnitudes and that the vector comprised of the remaining $n - k$ components has l_1 -norm less than some value, say Δ . We will refer to this type of signal as an approximately k -sparse signal, or for brevity only as an approximately sparse. The interested reader can find more on similar types of problems in [11].

In the rest of the paper we will further assume that the number of the measurements is $m = \delta n$ and the number of the “large” components of \mathbf{x} is $k = \zeta n$, where $0 < \zeta < 1$ and $0 < \delta < 1$ are constants independent of n (clearly, $\delta > \zeta$). This problem setup is more realistic of practical applications than the standard compressed sensing of ideally k -sparse signals (see, e.g., [18] and the references therein).

A particular way of solving (1) which has recently generated a large amount of research is called l_1 -optimization (basis pursuit) [2]. It proposes solving the following problem

$$\begin{aligned} \min \quad & \|\mathbf{x}\|_1 \\ \text{subject to} \quad & A\mathbf{x} = \mathbf{y}. \end{aligned} \quad (2)$$

Quite remarkably in [2] the authors were able to show that if the number of the measurements is $m = \delta n$ and if the matrix A satisfies a special property called the restricted isometry property (RIP), then any unknown vector \mathbf{x} with no more than $k = \zeta n$ (where ζ is an absolute constant which is a function of δ , but independent of n , and explicitly bounded in [2]) non-zero elements can be recovered by solving (2) (more on the case when the available measurements are noisy versions of \mathbf{y} can be found in e.g. [9], [10]). It turns out that for several specific classes of matrices, such as matrices with independent zero-mean Gaussian entries or independent Bernoulli entries, the RIP holds with overwhelming probability [2], [4], [5]. However, it should be noted that the RIP is only a sufficient condition for l_1 -optimization to produce a solution of (1). Instead of characterizing the $m \times n$ matrix A through the RIP condition, in [1], [3] the authors assume that A constitutes a k -neighborly polytope. It turns out (as shown in [1]) that this characterization of the matrix A is in fact a necessary and sufficient condition for (2) to produce the solution of (1). Furthermore, using the results of [6][14][15], it can be shown that if the matrix A has i.i.d. zero-mean Gaussian entries with overwhelming probability it also constitutes a k -neighborly polytope. The precise relation between m and k in order for this to happen is characterized in [1] as well. It should also be noted that for a given value m , i.e. for a given value of the constant δ , the value of the constant ζ is significantly better in [1], [3] than in [2].

As mentioned earlier, in this paper we will be interested in the case of approximately k -sparse signals. It is relatively easy to see that its exact recovery from a reduced number of measurements is not possible. Instead, we will pursue sufficient and necessary conditions on the matrix A to guarantee the closeness of the solution of (2) to \mathbf{x} for all or “almost all” possible approximately k -sparse signals. In particular, we determine necessary and sufficient conditions for the measure-

ment matrix A under which l_1 minimization guarantees the robustness of sparse signal recovery in the “weak”, “sectional” and “strong” senses as defined in the following sections. For example, in the strong sense of robustness, if the unknown approximately k -sparse vector is \mathbf{x} and $\hat{\mathbf{x}}$ is the solution of (2) then for any given constant $0 \leq \delta \leq 1$ there will be a constant ζ and measurement matrices A 's such that

$$\|\hat{\mathbf{x}} - \mathbf{x}\|_1 \leq \frac{2(C+1)\Delta}{C-1} \quad (3)$$

holds for all \mathbf{x} , where $C > 1$ is a given constant determining how close in l_1 norm the recovered vector $\hat{\mathbf{x}}$ should be to the original approximately k -sparse vector \mathbf{x} . As expected, ζ will be a function of C and δ . However, ζ will be an absolute constant independent of n . A similar problem was considered in [11], in which the performance bounds are obtained from the restricted isometry property from [2] and no explicit values of ζ were given. Since the RIP condition is only a sufficient condition, it generally gives rather loose bounds on the explicit values of ζ even in the ideally sparse case. Here we will compute sharp bounds on the explicit values of the allowable constants ζ for the general cases $C \geq 1$ in “weak”, “sectional” and “strong” robustness based on high-dimensional geometrical analysis for the null-space of A .

II. THE CONDITIONS ON A FOR ROBUSTNESS

We begin by introducing the three forms of robustness we will consider in this paper. Each is a stronger version of the one preceding it. In what follows, as before A is the $m \times n$ ($m < n$) measurement matrix, \mathbf{x} is the unknown $n \times 1$ vector, and $\mathbf{y} = A\mathbf{x}$. K denotes a subset of $\{1, 2, \dots, n\}$ with cardinality $|K| = k$ and $\bar{K} = \{1, 2, \dots, n\} \setminus K$.

- 1) Let K be fixed and \mathbf{x}_K , the value of \mathbf{x} on this set be also fixed. Then the solution produced by (2), $\hat{\mathbf{x}}$, will be called *weakly robust* if, for some $C > 1$ and all $\mathbf{x}_{\bar{K}}$, it holds that

$$\|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1} \|\mathbf{x}_{\bar{K}}\|_1$$

The above “weak” notion allows us to bound the error $\|\mathbf{x} - \hat{\mathbf{x}}\|_1$ in the following way. If the matrix A_K , obtained by retaining only those columns of A that are indexed by K , has full column rank, then the quantity

$$\kappa = \max_{A\mathbf{w}=0, \mathbf{w} \neq 0} \frac{\|\mathbf{w}_K\|_1}{\|\mathbf{w}_{\bar{K}}\|_1},$$

must be finite ($\kappa < \infty$). In particular, since $\mathbf{x} - \hat{\mathbf{x}}$ is in the null space of A ($\mathbf{y} = A\mathbf{x} = A\hat{\mathbf{x}}$), we have

$$\begin{aligned} \|\mathbf{x} - \hat{\mathbf{x}}\|_1 &= \|(\mathbf{x} - \hat{\mathbf{x}})_K\|_1 + \|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \\ &\leq (1 + \kappa) \|(\mathbf{x} - \hat{\mathbf{x}})_{\bar{K}}\|_1 \\ &\leq \frac{2C(1 + \kappa)}{C-1} \|\mathbf{x}_{\bar{K}}\|_1. \end{aligned}$$

Thus bounding the recovery error by the l_1 norm of the “approximately sparse” part of \mathbf{x} .

- 2) Let K be fixed. Then the solution produced by (2), $\hat{\mathbf{x}}$, will be called *sectionally robust* if, for some $C > 1$ and all $\mathbf{x} \in \mathbf{R}^n$, it holds that

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\bar{K}}\|_1.$$

Note that the above “sectional” notion requires that only the support set K be fixed. The unknown vector \mathbf{x} is otherwise arbitrary.

- 3) For a given $C > 1$ and integer k , the solution produced by (2), $\hat{\mathbf{x}}$, will be called *strongly robust* if, for all $K \subseteq \{1, 2, \dots, n\}$ of cardinality $|K| = k$ and all $\mathbf{x} \in \mathbf{R}^n$, it holds that

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq \frac{2(C+1)}{C-1} \|\mathbf{x}_{\bar{K}}\|_1.$$

The “strong” robust notion requires only that k be fixed. The support set and unknown vector are otherwise arbitrary.

We can now give necessary and sufficient conditions on the measurement matrix A to satisfy each of the above notions for the robustness of l_1 norm recovery.

Theorem 1: 1) For a given $C > 1$, support set K , and \mathbf{x}_K , the solution $\hat{\mathbf{x}}$ produced by (2) will be *weakly robust* if, and only if, $\forall \mathbf{w} \in \mathbf{R}^n$ such that $A\mathbf{w} = 0$, we have

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\bar{K}}}{C} \right\|_1 \geq \|\mathbf{x}_K\|_1; \quad (4)$$

- 2) For a given $C > 1$ and support set K , the solution $\hat{\mathbf{x}}$ produced by (2) will be *sectionally robust* if, and only if, $\forall \mathbf{x}' \in \mathbf{R}^n$ and $\forall \mathbf{w} \in \mathbf{R}^n$ such that $A\mathbf{w} = 0$,

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\bar{K}}}{C} \right\|_1 \geq \|\mathbf{x}'_K\|_1; \quad (5)$$

- 3) For a given $C > 1$, the solution $\hat{\mathbf{x}}$ produced by (2) will be *strongly robust* if and only if $\forall K \subseteq \{1, 2, \dots, n\}, \forall \mathbf{x}' \in \mathbf{R}^n, \forall \mathbf{w} \in \mathbf{R}^n$ such that $A\mathbf{w} = 0$, we have

$$\|\mathbf{x}'_K + \mathbf{w}_K\|_1 + \left\| \frac{\mathbf{w}_{\bar{K}}}{C} \right\|_1 \geq \|\mathbf{x}'_K\|_1. \quad (6)$$

Proof: For reasons of space, we will only prove the result for weak robustness. The other two are proven similarly.

Sufficiency: Let $\mathbf{w} = \hat{\mathbf{x}} - \mathbf{x}$, for which $A\mathbf{w} = A(\hat{\mathbf{x}} - \mathbf{x}) = 0$. Since $\hat{\mathbf{x}}$ is the minimum l_1 norm solution, we have $\|\mathbf{x}\|_1 \geq \|\hat{\mathbf{x}}\|_1 = \|\mathbf{x} + \mathbf{w}\|_1$, and therefore $\|\mathbf{x}_K\|_1 + \|\mathbf{x}_{\bar{K}}\|_1 \geq \|\hat{\mathbf{x}}_K\|_1 + \|\hat{\mathbf{x}}_{\bar{K}}\|_1$. Thus,

$$\begin{aligned} \|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1 &\geq \|\mathbf{w}_{\bar{K}} + \mathbf{x}_{\bar{K}}\|_1 - \|\mathbf{x}_{\bar{K}}\|_1 \\ &\geq \|\mathbf{w}_{\bar{K}}\|_1 - 2\|\mathbf{x}_{\bar{K}}\|_1. \end{aligned}$$

But the condition (4) guarantees that

$$\|\mathbf{w}_{\bar{K}}\|_1 \geq C(\|\mathbf{x}_K\|_1 - \|\mathbf{x}_K + \mathbf{w}_K\|_1),$$

so we have

$$\|\mathbf{w}_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|\mathbf{x}_{\bar{K}}\|_1,$$

and

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1} \|\mathbf{x}_{\bar{K}}\|_1,$$

as desired.

Necessity: Since in the above proof of the sufficiency, equalities can be achieved in the triangular inequalities, the conditions (4), (5) and (6) are also necessary conditions for the respective robustness to hold for every \mathbf{x} . (Otherwise, for certain \mathbf{x} 's, there will be $\mathbf{x}' = \mathbf{x} + \mathbf{w}$ with $\|\mathbf{x}'\|_1 < \|\mathbf{x}\|_1$ while violating the respective robustness definitions. Also, such \mathbf{x}' can be the solution to (2)). ■

We should remark that, after we take $C = 1$ and let (4), (5) and (6) take strict inequality for all $\mathbf{w} \neq 0$ in the null space of A , the conditions (4), (5) and (6) are also sufficient and necessary conditions for unique exact recovery of ideally k -sparse signals in “weak”, “sectional” and “strong” senses [1], namely the unique exact recovery of a specific ideally k -sparse signal, the unique exact recoveries of all ideally k -sparse signal on a specific support set K and the unique exact recoveries of all ideally k -sparse signal on all possible support sets K . In fact, if $\|\mathbf{x}_K\|_1 = 0$, from similar triangular inequality derivations in Theorem 1, we have $\hat{\mathbf{x}} = \mathbf{x}$ under all the three conditions. Please also note that related but different null-space characterizations appear in many other works, for example, [8], [19] and [20] etc. But the distinct contribution here is the establishment of the conditions for “weak”, “sectional” and “strong” robustness, and at the same time, we give *sharp* performance bounds for these null-space characterizations.

For a given value $\delta = \frac{m}{n}$ and any value $C \geq 1$, the question of interest is to determine the range of $\zeta = \frac{k}{n}$ for which the three notions of robust recovery are satisfied. As one may imagine, for a specific A , it is very hard to check whether any of these three conditions hold. Therefore one instead should focus on certain random ensembles of measurement matrices and should ask the question of for what ranges of δ and ζ do the three robust recovery conditions hold for a randomly chosen A from this ensemble with overwhelming probability as $n \rightarrow \infty$? The standard results on compressed sensing assume that the matrix A has iid $\mathcal{N}(0, 1)$ entries. In this case, the null-space of A has a rotationally invariantly distribution. Sampling from this rotationally invariant distribution is equivalent to uniformly sampling a random $(n - m)$ -dimensional subspace from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$, the set of $(n - m)$ -dimensional subspaces in the n -dimensional Euclidean space R^n [7]. For such a distribution on A and ideally sparse signals, the sharp bounds of [1], for example, apply. In this paper, we will use a Grassmann angle approach to analyze the probabilities P_1 , P_2 , and P_3 that (4), (5) and (6) *fail* respectively, thus giving performance bounds on δ for ζ to satisfy these three conditions respectively.

III. THE GRASSMANN ANGLE APPROACH FOR THE NULL-SPACE CHARACTERIZATIONS

Let us first focus on deriving the probability P_1 that the weak robustness condition fails under the sparsity $|K| = k$, namely the probability there exists a vector $\mathbf{w} \in Z$ failing the condition (4), if we uniformly sample a random $(n - m)$ -dimensional subspace Z from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. By scaling, we can restrict our attention to those

vectors \mathbf{x} that are only supported on K (or a subset of K) on the cross-polytope $\{\mathbf{x} \in R^n \mid \|\mathbf{x}\|_1 = 1\}$.

By symmetry, without loss of generality, we assume the signs of the elements of \mathbf{x} to be non-negative. Since \mathbf{x} is supported on the set K (or a subset of K) and is restricted to the cross-polytope $\{\mathbf{x} \in R^n \mid \|\mathbf{x}\|_1 = 1\}$, \mathbf{x} is also on a $(k - 1)$ -dimensional face, denoted by F , of the skewed cross-polytope SP:

$$\text{SP} = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\} \quad (7)$$

For the time being, we assume \mathbf{x} is in the relative interior of this $(k - 1)$ dimensional face F . For this particular \mathbf{x} on F , the probability P_1 , that $\exists \mathbf{w} \in Z$ ($\mathbf{w} \neq 0$) such that

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\bar{K}}}{C}\|_1 \leq \|\mathbf{x}_K\|_1 = 1, \quad (8)$$

is essentially the probability that a uniformly chosen $(n - m)$ dimensional subspace Z shifted by the point \mathbf{x} , namely $(Z + \mathbf{x})$, intersects the skewed cross-polytope SP *non-trivially*, namely, at some other point besides \mathbf{x} .

From the linear property of the subspace Z , the event that $(Z + \mathbf{x})$ intersects the skewed crosspolytope SP is equivalent to the event that Z intersects nontrivially with the cone $\text{SP-Cone}(\mathbf{x})$ obtained by observing the skewed polytope SP from the point \mathbf{x} . (Namely, $\text{SP-Cone}(\mathbf{x})$ is the conic hull of the point set $(\text{SP} - \mathbf{x})$ and of course $\text{SP-Cone}(\mathbf{x})$ has the origin of the coordinate system as its apex.) As noticed in the geometry for convex polytopes [12][13], the $\text{SP-Cone}(\mathbf{x})$ are identical for any \mathbf{x} lying in the relative interior of the face F . This means that the event that $(Z + \mathbf{x})$ intersects non-trivially with the skewed cross-polytope SP is the same for any relative interior point \mathbf{x} on the face F . Thus, the probability P_1 is the same for any such \mathbf{x} 's. (The acute reader may have noticed that $\mathbf{x} \in F$ may not be in the relative interior of F , but it turns out that the $\text{SP-Cone}(\mathbf{x})$ in this case is only a subset of the cone we get when \mathbf{x} is in the relative interior of F . So we do not lose anything if we restrict \mathbf{x} to be in the relative interior of the face F).

From its definition, P_1 is exactly the *complementary Grassmann angle* [12] for the face F with respect to the polytope SP under the Grassmann manifold $\text{Gr}_{(n-m)}(n)$: the probability of a uniformly distributed $(n - m)$ -dimensional subspace Z from the Grassmannian manifold $\text{Gr}_{(n-m)}(n)$ intersecting non-trivially with the cone $\text{SP-Cone}(\mathbf{x})$ formed by observing the skewed cross-polytope SP from the relative interior point \mathbf{x} . Building on the works by L.A.Santaló [16] and P.McMullen [17] etc. in high dimensional integral geometry and convex polytopes, the complementary Grassmann angle for the $(k - 1)$ -dimensional face F can be explicitly expressed as the sum of products of internal angles and external angles [13]:

$$P_1 = 2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}), \quad (9)$$

where s is any nonnegative integer, G is any $(m + 1 + 2s)$ -dimensional face of the skewed crosspolytope $(\mathfrak{S}_{m+1+2s}(\text{SP}))$ is the set of all such faces), $\beta(\cdot, \cdot)$ stands for the internal

angle and $\gamma(\cdot, \cdot)$ stands for the external angle. Here the internal angles and external angles are defined as [13], [17]:

- An internal angle $\beta(F_1, F_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1 .¹ The internal angle $\beta(F_1, F_2)$ is defined to be zero when $F_1 \not\subseteq F_2$ and is defined to be one if $F_1 = F_2$.
- An external angle $\gamma(F_3, F_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3 . The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \not\subseteq F_4$ and is defined to be one if $F_3 = F_4$.

Note that there are $\binom{n}{k}$ possible support sets K and there are 2^k possible sign patterns for signal \mathbf{x}_K . From previous discussions, we know that the event that the condition (4) fails is the same for all \mathbf{x}_K 's of a specific support set and a specific sign pattern. Then we can upper-bound P_2 and P_3 as

$$P_2 \leq 2^k \times P_1, \quad P_3 \leq \binom{n}{k} \times 2^k \times P_1. \quad (10)$$

Therefore to compute (9) and (10) all we need are formulas for the internal and external angles. These are given in the lemmas below.

Lemma 1: Suppose that F is a $(k-1)$ -dimensional face of the skewed cross-polytope

$$\text{SP} = \{\mathbf{y} \in R^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\bar{K}}}{C}\|_1 \leq 1\}$$

supported on the subset K with $|K| = k$. Then the internal angle $\beta(F, G)$ between the $(k-1)$ -dimensional face F and a $(l-1)$ -dimensional face G ($F \subseteq G$) is given by

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \quad (11)$$

where $V_i(S^i)$ denotes the i -th dimensional surface measure on the unit sphere S^i , while $V_i(\alpha', i)$ denotes the surface measure for regular spherical simplex with $(i+1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i+1)$ vertices. (11) is equal to $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m'-1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta) \quad (12)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda \quad (13)$$

Proof: Wlog, assume that F is a $(k-1)$ -dimensional face with k vertices as $e^p, 1 \leq p \leq k$, where e^p is the n -dimensional standard unit vector with the p -th element as '1'; and assume that the $(l-1)$ -dimensional face G be the convex hull of the l vertices: $e^p, 1 \leq p \leq k$ and $Ce^p, (k+1) \leq p \leq l$. Then the epicenter of the face F is $Ec = \sum_{p=1}^k e^p/k$.

¹Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

By symmetry, it is not hard to see that the cone formed by seeing another $(l-1)$ -dimensional face G of the skewed cross-polytope SP from an interior point x^F of the face F is the direct sum of the linear hull $\text{lin}\{F - x^F\}$ and a regular $(l-k)$ -dimensional positive hull of $(l-k)$ vectors a^1, a^2, \dots, a^{l-k} , where $a^i = C \times e^{k+i} - \sum_{p=1}^{p=k} e^p/k, 1 \leq i \leq (l-k)$. The normalized inner products between any two of these $(l-k)$ vectors is

$$\frac{\langle a^i, a^j \rangle}{\|a^i\| \|a^j\|} = \frac{k \times \frac{1}{k^2}}{C^2 + k \times \frac{1}{k^2}} = \frac{1}{1 + kC^2}. \quad (14)$$

By the computed volumes of such cones [6][15], we have the result stated in the lemma. ■

Lemma 2: Let F, G and SP be the same as in the previous lemma. Then $\gamma(G, \text{SP})$ between the $(l-1)$ -dimensional face G and the skewed cross-polytope SP is:

$$\gamma(G, \text{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^{\infty} e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx. \quad (15)$$

Proof: Wlog, consider the $(l-1)$ -dimensional face G as the convex hull $\text{conv}\{C \times e^l, \dots, C \times e^{k+1}, e^k, \dots, e^1\}$ of the skewed cross-polytope SP. The 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing G are given by $\{\sum_{p=l+1}^n j_p e^p/C + \sum_{p=k+1}^l e^p/C + \sum_{p=1}^k e^p, j_p \in \{-1, 1\}\}$. Then the outward normal cone $c(G, \text{SP})$ at the face G is the positive hull of these normal vectors. Thus

$$\int_{c(G, \text{SP})} e^{-\|x\|^2} dx = \gamma(G, \text{SP}) V_{n-l}(S^{n-l}) \times \int_0^{\infty} e^{-r^2} r^{n-l} dr = \gamma(G, \text{SP}) \cdot \pi^{(n-l+1)/2}, \quad (16)$$

where $V_{n-l}(S^{n-l})$ is the spherical volume of the $(n-l)$ -dimensional sphere S^{n-l} . Now let $U = \{x \in R^{n-l+1} \mid x_1 \geq 0, |x_p| \leq \frac{x_1}{C}, 2 \leq p \leq (n-l+1)\}$ and $f : U \rightarrow c(G, \text{SP})$ be the linear and bijective map $f(x_1, \dots, x_{n-l+1}) = \sum_{p=l+1}^n x_{p-l+1} e^p + \sum_{p=k+1}^l \frac{x_1}{C} e^p + \sum_{p=1}^k x_1 \times e^p$. Then,

$$\begin{aligned} \int_{c(G, \text{SP})} e^{-\|x'\|^2} dx' &= \sqrt{k + \frac{l-k}{C^2}} \int_U e^{-\|x\|^2} dx \\ &= \int_0^{\infty} \int_{-\frac{x_1}{C}}^{\frac{x_1}{C}} \dots \int_{-\frac{x_1}{C}}^{\frac{x_1}{C}} e^{-x_2^2 - \dots - x_{n-l+1}^2 - (k + \frac{l-k}{C^2}) x_1^2} \\ &\quad \sqrt{k + \frac{l-k}{C^2}} dx_2 \dots dx_{n-l+1} dx_1 \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_0^{\infty} e^{-(k + \frac{l-k}{C^2}) x^2} \left(\int_{-\frac{x}{C}}^{\frac{x}{C}} e^{-y^2} dy \right)^{n-l} dx \\ &= 2^{n-l} \int_0^{\infty} e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx, \end{aligned}$$

where $\sqrt{k + \frac{l-k}{C^2}}$ is due to the change of integral variables. Combining it with (16) leads to the desired result. ■

IV. NUMERICAL COMPUTATIONS ON THE BOUNDS

Now that we have computed the internal and external angles, we can in principle compute P_1 , P_2 and P_3 from (9) and (10). Of course, what we are interested in is whether P_i ($i = 1, 2, 3$) go to zero as $n \rightarrow \infty$. It turns out that (9) and (10) have exponents that are (dominantly) linear in n and so if the coefficient of the exponent is negative, $P_i \rightarrow 0$, as $n \rightarrow \infty$. Computing these exponents requires computing the exponents of the internal and external angles from Lemmas 1 and 2, respectively, (using the Laplace method, or otherwise) and computing the exponents of the combinatorial terms arising from the sums in (9). Doing all this (and omitting the details for lack of space) we arrive at the following result.

Theorem 2 (Recovery thresholds): Fix $\beta = \frac{m}{n}$ and $C > 1$ and assume that A is drawn from the iid Gaussian ensemble with $\mathcal{N}(0, 1)$ entries. Then, as $n \rightarrow \infty$, the l_1 minimization (2) will with overwhelming probability recover \mathbf{x} in the weakly robust sense for all $k \leq \zeta n$, if for all $\beta \leq \nu \leq 1$ it holds that

$$H\left(\frac{\nu - \zeta}{1 - \zeta}\right)(1 - \zeta) - \psi_{ext} - \psi_{int} < 0, \quad (17)$$

uniformly. Here $\psi_{ext} = -(1 - \nu) \log(\text{erf}(x_\nu)) + \nu x_\nu^2$, where x_ν is the unique solution to

$$\sqrt{8\pi} x \text{erf}(x) e^{\frac{x^2}{2}} = \frac{1 - \nu}{(C^2 - 1)\zeta + \nu}.$$

Furthermore, ψ_{int} is given by

$$(\nu - \zeta) \left[\log 2 + \frac{(\nu - \zeta)y_\nu^2}{2\zeta C^2} + \max_s (sy_\nu - s^2 - \log 2\Phi(s)) \right],$$

where y_ν is the unique minimizer of

$$\frac{(\nu - \zeta)y^2}{2\zeta C^2} + \max_s (sy - s^2 - \log 2\Phi(s)),$$

and $\Phi(\cdot)$ is the CDF of the zero-mean unit variance Gaussian distribution. For the sectional and strong robust recoveries the conditions respectively become

$$\nu \log(2) + H\left(\frac{\nu - \zeta}{1 - \zeta}\right)(1 - \zeta) - \psi_{ext} - \psi_{int} < 0,$$

$$\nu \log(2) + H(\zeta) + H\left(\frac{\nu - \zeta}{1 - \zeta}\right)(1 - \zeta) - \psi_{ext} - \psi_{int} < 0.$$

We can use the above theorem to explicitly compute the “weak”, “sectional” and “strong” thresholds on ζ , as a function of β and C . For comparison, following [1], we take $\beta = 0.5555$ and plot the thresholds as a function of C in Figure 1. When $C = 1$, we get the same thresholds for ζ as in [1] on k/n for ideally sparse signals. As expected, as C grows, l_1 minimization achieves higher signal recovery accuracy, and then also requires a smaller sparsity level ζ .

REFERENCES

[1] David Donoho, “High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension”, *Discrete and Computational Geometry*, 102(27), pp. 617-652, 2006, Springer.

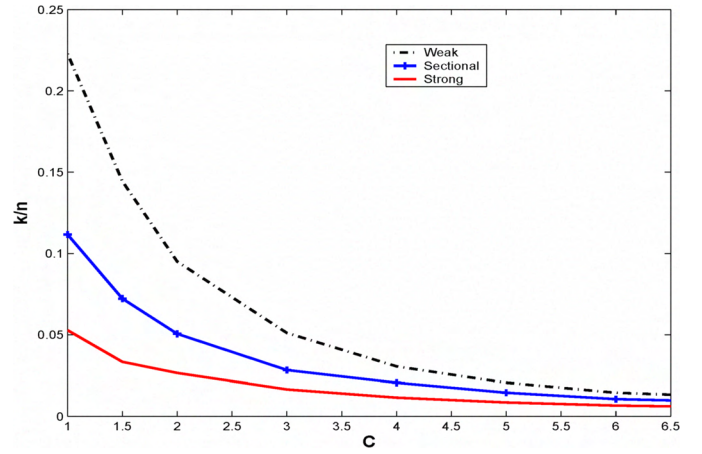


Fig. 1. Allowable sparsity for several robustness bounds as a function of C

[2] Emmanuel Candes and Terence Tao, “Decoding by linear programming”, *IEEE Trans. on Information Theory*, 51(12), pp. 4203 - 4215, December 2005.

[3] David Donoho and Jared Tanner, “Thresholds for the Recovery of Sparse Solutions via l_1 Minimization”, *Proceedings of the Conference on Information Sciences and Systems*, March 2006.

[4] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, “A simple proof of the restricted isometry property for random matrices”, *To appear in Constructive Approximation*, available online at <http://www.dsp.ece.rice.edu/cs/>.

[5] Mark Rudelson and Roman Vershynin, “Geometric approach to error correcting codes and reconstruction of signals”, *International Mathematical Research Notices*, 64, pp. 4019 - 4041, 2005.

[6] A.M. Vershik and P.V. Sporyshev, “Asymptotic Behavior of the Number of Faces of Random Polyhedra and the Neighborliness Problem”, *Selecta Mathematica Sovietica* vol. 11, No. 2 (1992).

[7] W. M. Boothby, *An Introduction to Differential Manifolds and Riemannian Geometry*, 2nd ed. San Diego, CA: Academic, 1986.

[8] Y. Zhang, “When is missing data recoverable”, available online at <http://www.dsp.ece.rice.edu/cs/>.

[9] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections”, *IEEE Trans. on Information Theory*, 52(9), pp. 4036-4048, September 2006.

[10] M. J. Wainwright, “Sharp thresholds for high-dimensional and noisy recovery of sparsity”, *Proc. Allerton Conference on Communication, Control, and Computing*, Monticello, IL, September 2006.

[11] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best k -term approximation”. (Preprint, 2006), available online at <http://www.dsp.ece.rice.edu/cs/>.

[12] Branko Grünbaum, Grassmann angles of convex polytopes. *Acta Math.*, 121:pp.293-302, 1968.

[13] Branko Grünbaum, *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler.

[14] Fernando Affentranger and Rolf Schneider, “Random projections of regular simplices”. *Discrete Comput. Geom.*, 7(3):pp.219-226, 1992.

[15] Károly Böröczky, Jr. and Martin Henk, “Random projections of regular polytopes”. *Arch. Math. (Basel)*, 73(6):pp.465-473, 1999.

[16] L.A. Santaló, *Geometría integral en espacios de curvatura constante*, Rep.Argentina Publ.Com.Nac.Energí Atómica, Ser.Mat 1,No.1(1952)

[17] Peter McMullen, “Non-linear angle-sum relations for polyhedral cones and polytopes”. *Math. Proc. Cambridge Philos. Soc.*, 78(2):pp.247-261, 1975.

[18] Emmanuel Candes, “Compressive sampling”, *Proc. International Congress of Mathematics*, 3, pp. 1433-1452, Madrid, Spain, 2006.

[19] Weiyu Xu and Babak Hassibi, “Compressive Sensing over the Grassmann Manifold: A Unified Analytical Approach,” *Allerton Conference on Communication, Control, and Computing*, September 2008

[20] Boris S. Kashin and Vladimir N. Temlyakov, “A remark on compressed sensing,” *Mathematical Notes*, 82(5-6), pp. 748-755, Nov. 2007.